

AD-A161 957

THE EFFECTS OF TRAINING ON RATERS' ACCURACY AND
COGNITIVE CATEGORIES(U) MICHIGAN STATE UNIV EAST
LANSING DEPT OF PSYCHOLOGY C OSTROFF ET AL. NOV 85

1/1

UNCLASSIFIED

TR-85-5 N00014-83-K-0756

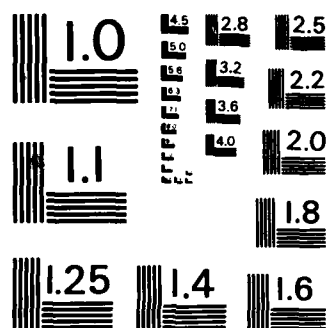
F/G 5/9

NL

END

FORMED

QTH



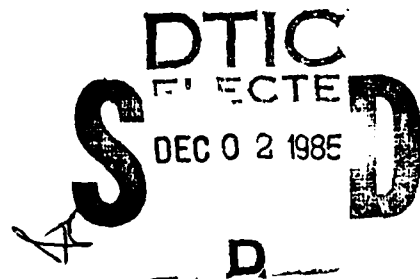
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICHIGAN STATE UNIVERSITY

Industrial/Organizational Psychology and Organizational Behavior

AD-A161 957

The Effects of Training on Raters' Accuracy
and Cognitive Categories
Cheri Ostroff and Daniel R. Ilgen
Michigan State University



DTIC FILE COPY

Michigan State University
East Lansing, Michigan 48824

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

85 11 25 132

12

The Effects of Training on Raters' Accuracy
and Cognitive Categories
Cheri Ostroff and Daniel R. Ilgen
Michigan State University

Prepared for
Office of Naval Research
Organizational Effectiveness Research Programs
Code 4420E

Grant No. N00014-83-K-0756
NR170-961

Technical Report No. 85-5
Department of Psychology
and
Department of Management
Michigan State University

UNCLASSIFIED

DTIC
SELECTED
DEC 02 1985
S D

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 85-5	2. GOVT ACCESSION NO. AN-A161 957	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Effects of Training on Raters' Accuracy and Cognitive Categories		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER 2010
7. AUTHOR(s) Cheri Ostroff and Daniel R. Ilgen		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0756
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Michigan State University East Lansing, MI 48824-1117		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR170-961
11. CONTROLLING OFFICE NAME AND ADDRESS Organizational Effectiveness Research Programs Office of Naval Research (Code 4420E) Arlington, VA 22217		12. REPORT DATE November, 1985
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Rater training, performance ratings, rating accuracy, cognitive categories		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This research investigated the effects of two forms of rater accuracy training on rating accuracy and raters' cognitive categories for appraising performance. Seventy-one hospital nurses completed measures designed to assess the cognitive categories they used to judge others. They also rated the performance of a nurse presented on videotape. Raters then received one of two training programs and again viewed the videotape, rated the performance of the videotaped ratee, and completed the measures assessing cognitive categories. Both training programs improved rating accuracy and the match between the raters' cognitive categories.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

and the dimensions used to assess performance. Training based on personalized feedback of raters' categories tended to yield more accurate ratings and led to greater changes in raters' cognitive categories.

Accession	
1. Title	
2. Author	
3. Subject	
4. Date	
5. Place	
6. Price	
7. Notes	
8. Remarks	
9. Indexing	
10. Classification	
11. Status	
12. Remarks	
13. Remarks	
14. Remarks	
15. Remarks	
16. Remarks	
17. Remarks	
18. Remarks	
19. Remarks	
20. Remarks	
21. Remarks	
22. Remarks	
23. Remarks	
24. Remarks	
25. Remarks	
26. Remarks	
27. Remarks	
28. Remarks	
29. Remarks	
30. Remarks	
31. Remarks	
32. Remarks	
33. Remarks	
34. Remarks	
35. Remarks	
36. Remarks	
37. Remarks	
38. Remarks	
39. Remarks	
40. Remarks	
41. Remarks	
42. Remarks	
43. Remarks	
44. Remarks	
45. Remarks	
46. Remarks	
47. Remarks	
48. Remarks	
49. Remarks	
50. Remarks	
51. Remarks	
52. Remarks	
53. Remarks	
54. Remarks	
55. Remarks	
56. Remarks	
57. Remarks	
58. Remarks	
59. Remarks	
60. Remarks	
61. Remarks	
62. Remarks	
63. Remarks	
64. Remarks	
65. Remarks	
66. Remarks	
67. Remarks	
68. Remarks	
69. Remarks	
70. Remarks	
71. Remarks	
72. Remarks	
73. Remarks	
74. Remarks	
75. Remarks	
76. Remarks	
77. Remarks	
78. Remarks	
79. Remarks	
80. Remarks	
81. Remarks	
82. Remarks	
83. Remarks	
84. Remarks	
85. Remarks	
86. Remarks	
87. Remarks	
88. Remarks	
89. Remarks	
90. Remarks	
91. Remarks	
92. Remarks	
93. Remarks	
94. Remarks	
95. Remarks	
96. Remarks	
97. Remarks	
98. Remarks	
99. Remarks	
100. Remarks	

QUALITY
INSPECTED
3

S/N 0102- LR-014-6601

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

The Effects of Training on Raters' Accuracy and Cognitive Categories

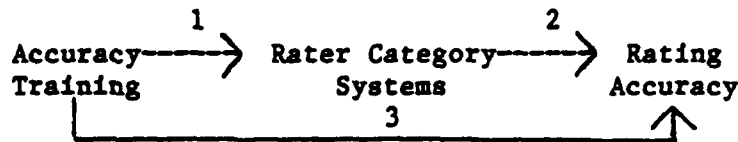
Recently, research on performance appraisal has shifted from a focus on rating formats and rating errors to rating process variables and rating accuracy. Considerable attention has been devoted to explicating the cognitive processes of raters involved in performance evaluation (Landy & Farr, 1980; Feldman, 1981; Ilgen & Feldman, 1983). The rating task is conceptualized as gathering, storing and recalling information. A central component in this process is the storing of information in cognitive categories or "bins" which guide attention to information about ratees and form the basis for recall of that information. To the extent that the rater's category system facilitates attention to storage and recall of relevant ratee behaviors, performance evaluations should be more accurate (Ilgen & Feldman, 1983). Ostroff and Ilgen (1985) demonstrated that these cognitive categories of raters do influence rating accuracy. This suggests that one way to improve performance ratings is to direct attention to improving the cognitive categories of raters.

Training programs aimed at increasing rater accuracy have typically employed two types of training--"error" training and "accuracy" training. Although attempts to train raters to avoid common psychometric errors were successful in doing so (Bernardin & Walter, 1977; Borman, 1975; Ivancevich, 1979; Latham, Wexley, & Pursell, 1975), subsequent studies demonstrated that accuracy was

relatively unaffected by the reduction of psychometric errors in ratings (Bernardin & Pence, 1980; Borman, 1975, 1979; Pulakos, 1984). To correct this problem, an alternative approach to error training was proposed by Bernardin and Buckley (1981) and Borman (1979) in which raters are trained to use a common frame-of-reference to assess ratee behaviors. Studies utilizing this "accuracy" approach typically concentrate on the a) multidimensionality of performance, b) specification of rating scale dimensions and the behaviors comprising these dimensions, and c) recognizing possible discrepancies between "true" ratings and the rater trainee ratings (cf. Bernardin & Pence, 1980; McIntyre, Smith & Haslett, 1984; Pulakos, 1984, in press). All studies found that rating accuracy improves with such training.

Two sets of indirect evidence suggest that accuracy training alters the cognitive dimensions people use to assess performance and that these new cognitive dimensions lead to more accurate ratings. First, accuracy training, which focuses on describing the performance dimensions used on the rating scale, increases accuracy presumably by bringing the rater's categories for judging performance in line with the scale dimensions. Whether the categories used by the rater are actually more consistent with the scale after training has not been tested. Second, those who have naturally occurring categories which are consistent with the scale tend to be more accurate than those whose categories are more discrepant from the scale (Ostroff & Ilgen, 1985). Taken together,

indirect evidence is provided for the sequence of training affecting categories which affect accuracy. The data support links two and three in the following sequence.



One purpose of the present study was to develop a rater training program directed at the categories raters used in evaluating performance. In this program, labeled feedback training program, raters' own cognitive categories were assessed, feedback about the match between raters' own categories and the categories relevant to the job was provided, and the effects of a good match of categories to rating scales was discussed. Specifically, raters were given feedback as to how well their categories matched rating scale dimensions, the extent to which they distinguished between job relevant and irrelevant behaviors and dimensions, and the degree to which they differentiated among dimensions. It was believed that this more direct approach to linking the rater's categories to performance appraisals in a training program would be more useful in enhancing rater accuracy than the more general accuracy training approaches previously used. The following hypotheses were tested:

Hypothesis 1A: Providing raters with standard rater accuracy training or with feedback training will improve rater accuracy.

Hypothesis 2A: Raters receiving feedback training will show a greater improvement in rating accuracy than those provided with standard rater accuracy training.

A second purpose of this study was to determine if training does indeed affect the rater's cognitive categories as assumed. Although previous rater training programs assumed that the cognitive categories of raters are important and are affected by training, no research has addressed this issue directly. To the degree that the training program focuses directly on the rater's own cognitive categories, the training should have a greater influence on the rater's cognitive categories than a more general training program. This implies that:

Hypothesis 2A: Both rater accuracy training and feedback training will affect raters' cognitive categories.

Hypothesis 2B: Providing raters with feedback training will affect raters' cognitive categories to a greater extent than the receipt of accuracy training without feedback.

Method

Overview

The research required nurses' participation in three separate phases. The first phase involved an orientation session where nurses completed the pre-training research measures. In the second phase, nurses participated in one of two training programs. Both Phase One and Phase Two were conducted at the hospitals where the nurses were employed. Finally, in Phase Three, nurses responded by

mail to the post-training research measures.

Sample

Participants in Phase One were 125 nurses (97% female) from three large midwestern hospitals. Of these, 71 also participated in Phase Two and 53 participated in all three phases. From the total sample, 92% of the nurses had five or more years of work experience and 87% had previous experience rating nurses. Ninety-seven percent of the persons were in some type of supervisory position.

Stimulus Materials and Measures

Two sets of measures were used in the study. The Behavior Grid assessed raters' cognitive categories. To assess rating accuracy, a videotape of a nurse performing job duties, a performance rating scale and true score ratings were used. (For a more complete description of the development and reliability of these measures, see Ostroff & Ilgen, 1985).

Behavior Grid. The Behavior Grid was a matrix which contained brief descriptions of behaviors as rows (e.g., "would expect this nurse to give only a partial bath to an acutely ill cardiac patient in an oxygen tent") and performance dimensions as columns (e.g., "Knowledge and Judgment"). A brief definition of each dimension was also provided. In addition, one-half of the rows were behaviors judged to be relevant to job performance and the other half irrelevant. The same distinction was made for the dimensions (columns). For example, an irrelevant behavior was "this nurse

wears a lot of make-up, perfume or cologne to work" and an irrelevant dimension was "Appearance."

The grid itself was composed of empty cells with each row and column labeled as described above and with job relevant and irrelevant rows and columns randomly ordered. To complete the grid, a nurse read each behavioral label for the row, then placed an "X" under the dimension column or columns that he or she felt the behavior represented. Given the examples mentioned above, a correct placement of the "giving a partial bath" behavior was under the "Knowledge and Judgment" dimension.

Cognitive Measures. From the Behavior Grid, the following six measures were derived:

1. Rating Scale Match

From the subset of dimensions on the Grid which were identified a priori as relevant to the nurse's job and a subset of behaviors which described those dimensions, an index of the match between the job and the rater's perception of dimensions and behaviors was derived. Each job relevant behavior received a score ranging from 6 to 1, depending on the degree of match to the rating scale. The highest score was given when an "X" appeared in the appropriate column for the behavior and in no other columns. The next highest score was for an "X" placed in the appropriate column and also in one other column. The scores continued to decrease in a similar fashion,

depending upon the nature of the response. The scores for each behavior were totalled so that the Rating Match scores ranged from 20 to 120. Higher scores indicated a greater match.

2. Non-Job Relevant Behavior Classification

This index was the sum of the number of times behaviors identified as non-job relevant were misclassified as belonging to job related dimensions.

3. Job Relevant Behavior Classification

In a manner similar to 2 above, the number of times job relevant behaviors were sorted into non-job relevant dimensions was tallied.

4. Overall Cognitive Differentiation

This index was computed by totalling the number of check marks (or number of times behaviors were placed in dimensions) each rater placed in the grid. Lower scores indicated greater differentiation.

5. Job Behavior Cognitive Differentiation

This index was computed in a manner similar to 4 above, but only for the job related behaviors in the grid.

6. Non-Job Cognitive Differentiation

In a manner similar to 4 above, the number of check marks each rater placed in the grid for non-job related behaviors was tallied.

Videotape and Rating Scale. A 25 minute videotape featuring a nurse in a hospital setting served as the stimulus material for ratings. It featured 18 one to three minute scenes depicting enactments of job behaviors from one or more of the five performance dimensions to be rated on the performance evaluation scale. Within each dimension, the ratee's behavior was designed to be consistent in effectiveness level, but across job dimensions, the effectiveness level varied. True scores were also generated from expert raters. These ratings served as the standard to which subjects' ratings were compared and from which performance accuracy indices were computed.

Ratings of the nurse's performance were made using Smith and Kendall's (1963) behaviorally anchored rating scale (BARS) developed specifically for hospital nurses. The five dimensions on the BARS were Knowledge and Judgment, Organizational Ability, Skill in Human Relations, Conscientiousness and Observational Ability.

Accuracy Measures. Two accuracy measures were calculated for each rater and served as dependent variables. Cronbach's (1955) component of overall accuracy was computed by squaring the difference between the rated and true scores and summing over all dimensions. Lower overall accuracy scores indicated greater accuracy. Correlational accuracy was also computed for each rater by correlating the true scores and the observed scores. Higher correlational accuracy scores indicated greater accuracy in terms

of the pattern of performance levels across dimensions for the rates.

Procedure

For the first phase, pre-training, nurses participated in a one and one-half hour long session and were assessed in groups of three to fifty persons per session. After a brief description of the project, nurses completed several questionnaires.¹ Next, they completed the Behavior Grid. Once completed, the questionnaires were collected. This was followed by an explanation of the rating scales and the videotape. Nurses then viewed the videotape and rated the performance of the videotaped nurse on the BARS scale.

Phase Two, training, was conducted approximately 4-6 weeks later. Raters again participated in a one and one-half hour long session in which they received one of the treatment (training) programs. Hospitals were randomly assigned to treatment groups. Immediately following training, nurses again observed and rated the videotaped nurse using the BARS scale.

Approximately 4-6 weeks following the training phase, nurses were mailed the Behavior Grid and were again asked to complete the grid following the procedure described in Phase One. Seventy-seven percent of the nurses, who had volunteered during Phase Two to complete the final questionnaire, returned the completed measure. The third phase will be referred to as the post-training session. In sum, accuracy was assessed during the first and second phases

and the cognitive measures were assessed during the first and third.

Manipulations

Two types of rater training programs were employed in the study--rater accuracy training and feedback training. All nurses received a brief explanation of the session, namely a description of its goal to increase the accuracy of their evaluations of others' performance. A brief description of the rating process as one of coding, storing and recalling information about ratees was presented emphasizing the storing of information in categories. The pre-training session, which included the same rating task, served as the no-training control.

Rater Accuracy Training. Accuracy training was designed to provide raters with a common frame-of-reference for considering ratee performance. This session was based, in part, on the procedure outlined by Pulakos (1984) for accuracy training. Nurses were first lectured on the multidimensionality of the job and on the importance of attending to performance related to these job dimensions. Participants were then given the BARS rating form. Global definitions of each dimension were given by the trainer, followed by an indepth description of the behaviors comprising the dimensions. The types of behaviors indicative of various effectiveness levels within each dimension were discussed by pointing out differences in the effective versus ineffective behaviors which served as scale anchors. Nurses then practiced

using the scales by rating a short sample of three, one to three minute, videotaped scenes.

Next, a random sample of nurses' ratings were placed on an overhead, by dimension, and the group discussed what particular ratee behaviors led them to their ratings. The trainer also provided feedback on the accuracy of their practice ratings. Next, common rating errors (halo, central tendency, leniency, contrast, first impression, similar-to-me and stereotype) were explained and the trainer pointed out examples of such errors in the practice ratings. Finally, participants viewed a second sample of videotape and again practiced making ratings. Group discussions and feedback on their accuracy followed as described for the first practice sample. An overview of the session ended the training program.

Feedback Training. The feedback training session was designed to incorporate specific feedback to raters on the cognitive categories they had used in evaluating others in the earlier session. The importance of a) focusing on specific behaviors rather than general traits, b) distinguishing between job relevant and non-job relevant behaviors, c) defining the appropriate behaviors for particular job dimensions, and d) differentiating between job dimensions were highlighted in the lecture. Each rater received a feedback form with scores derived from the Behavior Grid which they completed during Phase One. The trainer first explained the distinction between job behaviors and more general personal characteristics of ratees, following which an explanation of job

versus non-job related behaviors and dimensions ensued. Raters were directed to their feedback form to determine the percent of non-job related behaviors they had perceived as belonging to each of the five job dimensions (from the BARS) and the percent of job relevant behaviors they placed in dimensions irrelevant to job performance. It was explained to trainees that misclassifying the behaviors in such a way could lead to erroneous judgments of performance. Discussion then focused on job relevant behaviors and the importance of observing and defining the appropriate behaviors for each of the five performance dimensions. Thus, feedback was provided, for each job dimension, as to the percent of job behaviors correctly placed in the appropriate dimension. The trainer also explained the importance of differentiating between job dimensions as opposed to viewing every job behavior as belonging to every dimension. Raters then received feedback as to whether they were low, average or high in differentiating among dimensions and were told that if they received low or average scores, they need to focus on distinguishing between which behaviors belong in which dimensions. Raters were also instructed to pay particular attention, in the remainder of the session, to those dimensions for which they received low scores. Following the feedback discussion, the trainer provided accuracy and error training by following the procedure described above for accuracy training. However, in order to keep the length of the two training programs the same, only one practice and subsequent discussion,

rather than the two given in the accuracy session, was given to raters in feedback training.

Experimental Design

The hypotheses were tested with 2 x 2 factorial designs with repeated measures on the second factor. The first factor was Training (Accuracy versus Feedback). The repeated measure factor was two levels of training Experience (low experience for pre-training and high experience for post-training).

Results

Training Effects on Accuracy

The means and standard deviations of the two accuracy measures, overall and correlational accuracy are reported in Table 1. Within experience level, the two accuracy measures were highly intercorrelated ($r = .68$ for low experience and $r = .76$ for high experience).

A 2 x 2 (training x experience) ANOVA with repeated measures on the second factor was performed to assess training and experience effects on overall accuracy. Results indicated no significant main effect for training, $F(1,69)=.00$, $p=1.0$, but a significant main effect resulted for experience, $F(1,69)=17.14$, $p=.0001$. Mean comparisons using Newman-Keuls tests revealed that raters were more accurate, measured by overall accuracy, after high experience than low experience. No training x experience interactions were found, $F(1,69)=.08$, $p=.78$.

Table 1

Means and Standard Deviations of Accuracy Measures by Experience
and Training

<u>Experience</u>	<u>Overall Accuracy</u>			<u>Correlational Accuracy</u>		
	<u>Accuracy</u>	<u>Feedback</u>	<u>Total</u>	<u>Accuracy</u>	<u>Feedback</u>	<u>Total</u>
Low						
M	.77	.75	.76	.69	.73	.71
SD	.74	.50	.63	.22	.11	.18
N	38	33	71	37	33	70
High						
M	.44	.46	.45	.73	.75	.74
SD	.32	.34	.32	.10	.06	.09
N	38	33	71	37	33	70
Totals						
M	.61	.60	.60	.71	.74	.72
SD	.59	.45	.28	.17	.09	.14
N	76	66	142	74	66	140

With respect to Hypothesis One, recall that the pre-training accuracy scores (assessed in the first session, i.e. low experience) were derived after raters were given some orientation to appraisal processes. During the low experience, pre-training session, raters may have gained some information, on their own, simply as a result of experience with the task, that served to enhance their rating accuracy. If so, the training sessions may have enhanced rating accuracy for both types of training, but may not have supplemented this information enough to reflect differences in the training programs. That is, the experience provided by the first exposure to the task may have provided a base upon which improvements occurred for the second session, but there may not have been sufficient room for improvement to detect differences between treatments.

To determine if participation in both low and high experience sessions affected accuracy scores in the training programs differently, we identified those who had not been present at the first phase of research but had attended the training session ($n=24$). This group was labeled the low participation group in a 2×2 factorial design with two levels of participation and two levels of training, using only data from Phase Two. The means and standard deviations for the accuracy measures by participation are presented in Table 2. For overall accuracy scores, results indicated no significant main effect for training, $F(1,91)=.46$, $p=.5$. A significant main effect was found for participation,

$F(1,91)=4.10$, $p=.05$, and there was a trend toward a training x participation interaction $F(1,91)=3.16$, $p=.08$. Mean comparisons using Newman-Keuls tests revealed that raters who received Accuracy training and who participated only in the training session were less accurate than those raters in any other group.

All of the above analyses were performed for the correlational accuracy index, however no significant results were revealed. This was probably due to the fact that there was little variance in the correlational accuracy scores across the sample.

Training Effects on Cognitive Categories

The cognitive variables measured in the first and third phase were intercorrelated and are reported in Table 3. Within session, there were high intercorrelations between the cognitive measures. The low correlations across sessions (r 's ranged from .00 to .14) implied that the raters' scores on the cognitive measures changed across over time.

To test for training effects on raters' cognitive categories, a 2×2 (Training x Experience) multivariate analysis of variance with training as the fixed factor and experience as the repeated measure was performed including all six cognitive measures as dependent variables. Results of this MANOVA revealed no significant main effect for training, $F(5,47)=.36$, $p=.87$, but a significant main effect for experience, $F(5,47)=13.32$, $p=.001$, and a significant training x experience interaction, $F(5,47)=3.08$, $p=.02$. Due to the significant main and interaction effects, 2×2

Table 2

Means and Standard Deviations of Accuracy Measures by Participation and Training

<u>Participation</u>	<u>Overall Accuracy</u>			<u>Correlational Accuracy</u>		
	<u>Accuracy</u>	<u>Feedback</u>	<u>Total</u>	<u>Accuracy</u>	<u>Feedback</u>	<u>Total</u>
Low						
M	.76	.39	.67	.68	.77	.71
SD	.16	.13	.13	.05	.01	.04
N	18	6	24	18	6	24
High						
M	.44	.46	.45	.73	.75	.74
SD	.32	.34	.32	.10	.06	.09
N	38	33	71	37	33	70
Totals						
M	.54	.45	.50	.71	.75	.73
SD	.48	.33	.43	.14	.06	.11
N	56	39	95	55	39	94

Note. Low participation indicates attendance at only Phase Two. High participation is attendance at both Phase One and Two.

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Table 3

Intercorrelations of Cognitive Measures Obtained from Nurses Who
Attended All Three Phases

<u>Cognitive Measure</u>	<u>Pre-Training</u>					
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
1. Rating Scale Match	--					
Behavior Classification						
2. Non-Job	-.53					
3. Job	-.50	.63				
Cognitive Differentiation						
4. Overall	-.65	.84	.72			
5. Job	-.67	.80	.74	.96		
6. Non-Job	-.56	.80	.61	.94	.81	--
<u>Cognitive Measure</u>	<u>Post-Training</u>					
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
1. Rating Scale Match	--					
Behavior Classification						
2. Non-Job	-.36					
3. Job	-.24	.45				
Cognitive Differentiation						
4. Overall	-.34	.73	.61			
5. Job	-.41	.62	.60	.97		
6. Non-Job	-.18	.80	.54	.90	.75	--

(training x experience) ANOVAs with repeated measures on the second factor were performed for each of the six dependent variables to determine which of the cognitive measures contributed to the significant effects.

The means and standard deviations for the cognitive measures as affected by training and experience are presented in Table 4 and the results of the ANOVAs for the six cognitive measures appear in Table 5. As can be seen from Table 5, the primary effect on cognitive variables resulted from increases in exposure to the rating task and training from the first to the third phase. All six measures changed significantly and this change accounted for an average of 11% of the variance based upon the mean of the Omega squares for experience across the six variables. Inspection of the patterns of the means for the marginally significant interactions indicated that these did not alter the nature of the changes resulting from experience. In all cases, the shifts were toward improvement in the cognitive responses.

The interaction data were less clear cut. It was hypothesized that experience would lead to changes in the cognitive variables, but that those who received feedback training would change (improve) to a greater extent, after training, than those who received accuracy training. That is, an interaction effect was predicted such that no differences between groups would exist during the first phase, improvement would occur for both groups between Phase One and Phase Three assessments, and improvement in

Table 4

Means and Standard Deviations for Cognitive Variables by Training and Experience

<u>Cognitive Variable</u>	<u>Training x Experience</u>						<u>Experience</u>	
	<u>Accuracy (n=30)</u>			<u>Feedback (n=23)</u>			<u>Totals (n=53)</u>	
	<u>Low</u>	<u>High</u>	<u>Total</u>	<u>Low</u>	<u>High</u>	<u>Total</u>	<u>Low</u>	<u>High</u>
Rating Scale Match								
\bar{X}	87.27	105.50	96.37	89.35	105.70	97.54	88.17	105.60
SD	16.57	8.01	15.83	14.24	9.78	14.65	15.49	8.73
Behavior Classification								
Non-Job								
\bar{X}	4.50	2.97	3.73	5.61	1.39	3.50	4.98	2.28
SD	4.75	2.97	4.00	4.10	2.13	3.87	4.47	2.73
Job								
\bar{X}	2.33	0.67	1.50	2.83	0.57	1.70	2.55	0.62
SD	2.62	0.92	2.12	3.93	0.66	3.01	3.22	0.81
Cognitive Differentiation								
Overall								
\bar{X}	73.80	53.97	63.88	82.17	48.00	65.09	77.43	51.38
SD	38.62	12.00	30.07	27.88	8.97	26.79	34.33	11.10
Job								
\bar{X}	42.70	29.70	36.20	49.39	25.48	37.43	45.60	27.87
SD	20.88	8.07	17.01	19.59	5.81	18.71	20.41	7.42
Non-Job								
\bar{X}	31.10	24.27	27.68	32.78	22.52	27.65	31.83	23.51
SD	19.23	4.77	14.31	9.30	3.73	8.71	15.61	4.39

Table 5

Results of Analyses of Variance for Cognitive Measures

Dependent Variable	Rating Scale Match						Behavior Classification					
	Non-Job			Job			Non-Job			Job		
	df	F	p	ω^2	F	p	df	F	p	ω^2	F	p
Training	1	.20	.66	.00	.10	.75		.17	.68	.00		
Training x S's within groups		(178.45)			(13.74)			(5.80)				
Experience	1	54.53	.00	.01	16.18	.00		18.52	.00	.35		
Experience x Training	51	0.15	.70	.00	3.52	.07		0.42	.52	.00		
Experience x S's within groups	53	(142.83)			(13.31)			(5.42)				

(table continued)

Table 5 (cont.)

Results of Analyses of Variance for Cognitive Measures

Dependent Variable	Cognitive Differentiation							
	Overall				Job			
	F	P	ω^2		F	P	ω^2	Non-job
Effect								
Training	.06	.81	.00		.16	.70	.00	.99 .00
Training x S's within groups	(663.01)			(254.74)			(128.94)	
Experience	29.80	.00	.04	42.17	.00	.06	13.83	.00 .02
Experience x Training	2.10	.15	.00	3.69	.06	.00	0.56	.46 .00
Experience x S's within groups	(637.05)			(210.31)			(137.54)	

Note. Numbers in parentheses are the mean square error associated with the F test directly above them in the table. Omega squares were computed separately for each effect and thus are not directly comparable.

categories would be greater for those who received feedback training as compared to accuracy training. As reported earlier, a significant interaction was found when all six cognitive variables were used in the MANOVA. However, the univariate analyses revealed only marginally significant interaction effects for two of the six variables.

Given the significant interaction for the MANOVA and our interest in comparisons between training conditions, comparisons between training programs were conducted on the Phase Three data only. Three of the six comparisons were significant (Non-Job Behavior Classification, $F(1,52) = 4.65$, $p = .04$; Overall Cognitive Differentiation, $F(1,52) = 3.97$, $p = .05$; Job Cognitive Differentiation, $F(1,52) = 4.50$, $p = .04$). In these cases, training with individualized feedback and a discussion of cognitive categories in rating created more beneficial responses on cognitive variables than did accuracy training alone (Omega squares for the three variables were .06, .05, .06, respectively).

One additional descriptive feature of the cognitive data deserves mention. For all dependent variables, the variances of the scores were less in Phase Three than Phase One and for five of the six variables, the variances were less with feedback training than accuracy training (see Table 4). The pattern of variances provides some additional evidence for the positive effect of training experience on raters' cognitive categories and for the advantage of providing specific feedback.²

Discussion

Although previous research on rater accuracy training has demonstrated that such training can lead to more accurate appraisals (Bernardin & Pence, 1980; McIntyre, et al., 1984; Pulakos, 1984, in press), the research has provided little information about how accuracy training actually improves rating accuracy. The lack of such information limits the usefulness of the training due to a paucity of data on the factors that influence accuracy. Greater knowledge about what and how variables impact accuracy would provide guidance for the development of future accuracy training.

One of the most prevalent explanations for how accuracy training affects ratings is through its effect on the way in which raters organize and store information about ratees in memory. Yet, this explanation has been based primarily on indirect inferences from the social cognition literature, rather than from research directly addressing performance appraisals (Ilgen & Favero, 1985; Ostroff & Ilgen, 1985). The present study provides more direct support for the influence of training which is directed toward improving rating accuracy on the cognitive categories used by raters.

This study replicated previous findings that rater accuracy training actually improves the accuracy with which raters evaluate others' performance. Two types of training programs were utilized--standard accuracy training and feedback training which

incorporated giving personalized feedback on categories and a discussion of the effect of categories in the rating task. Both training programs significantly increased rating accuracy from pre to post-training experience.

Due to the fact that raters evaluated the ratee's performance twice, once during pre-training and once post-training, it was possible that experience or practice in rating, rather than training, was the factor leading to increased accuracy scores. While this explanation is plausible, there is some data to counter this argument. Raters who did not have previous pre-training experience, but received feedback training, had mean accuracy scores which were similar to those who attended both sessions. If experience alone was the explanation, then raters without prior experience should have had accuracy scores which were less than those with experience. This was not the case for those given feedback training only, but interestingly, this did occur for those raters who received only accuracy training.

Raters who received only accuracy training had accuracy scores which were less than the post-training accuracy scores of raters who participated in both sessions, and their mean accuracy scores were similar to those of the pre-training accuracy scores. Further, for those who participated in only the training session, those who received feedback training were more accurate than those who received accuracy training. Taken together, these results imply that feedback training has a stronger effect on accuracy

scores, but this may be moderated by the amount of time, or experience, spent in training. Raters who attended both sessions spent a total of three hours in training, but only one-half hour of time in these sessions reflected differences between accuracy and feedback training. For those who only attended the training session, one-half out of one and one-half hours differentiated feedback from accuracy training. Here, the ratio of time spent emphasizing feedback, rather than accuracy, training was greater and thus the differences in the training programs may not have been swamped by the other information presented. Our suggestion is that when only a one-time training program is implemented, feedback training appears to be a superior strategy when discussion of the effects of categories on the rating task is incorporated in training.

One question which arises concerns what component of the feedback training led to increases in accuracy. Originally, we believed that providing raters with personalized feedback about their category systems would serve to increase rating accuracy. However, those raters who did not participate in the pre-training session, but attended the feedback training session, did not receive personalized feedback about their categories (the feedback was derived from measures completed during the first session). Yet, even in the absence of personalized feedback, these raters were equally as accurate as those who attended both sessions, and were more accurate than those who only received accuracy training.

Therefore, it is likely that the discussion which focused specifically on the effects of categories in the rating process led to the increased accuracy, rather than the feedback per se. This discussion may have provided valuable information to raters about the rating task.

One of the unique aspects of this study was the investigation of and finding that training does influence categories, since studies using the accuracy approach have implicitly assumed this link. Based on the findings here, it is apparent that training did, in fact, influence raters' cognitive categories when assessments of the categories were made approximately one month prior to and one month after training. Both training programs had a positive effect on raters' categories. Further, for some of the cognitive category indices, Feedback training, which focused on categories, had a greater effect than Accuracy training. Training programs which directly focus on cognitive processes by providing individualized feedback to raters about their own category systems and by discussing the role of categories in the rating process may make it easier for them to identify and alter their idiosyncratic category systems and thus have a greater impact on the categories.

Although the effects of any of the variables in the study on appraisal accuracy were not very strong, they were relatively consistent with much of the accuracy research that uses the experimental design used here--that is, one in which raters view

videotapes with known standards of behaviors. Videotapes can be used for the standards of performance only if there is high agreement among expert judges about the behavior displayed on the tape. Without high agreement, the standard for judging accuracy is not well defined. The result of this high agreement is that the final stimulus tape may contain obvious behaviors which allows naive subjects to be quite accurate in their judgments. This was the case in the present study and we suspect in many other studies employing this paradigm. We would expect that our findings with respect to accuracy and the findings of others may be conservative. That is, in job settings with more abstract behaviors, the effects should be stronger. However, there is a need to seek other paradigms for accuracy research in order to replicate these and other findings on accuracy.

The model used for rater training assumes that training affects cognitive categories which, in turn, affect accuracy. Results discussed thus far have indicated that training affects categories and training affects accuracy. Categories of raters have also been shown to be related to accuracy, but these measures were derived prior to training (Ostroff & Ilgen, 1985). Thus, additional correlational analyses were performed to determine if post-training categories of raters were related to rating accuracy after training. Non-Job Cognitive Differentiation and Rating Scale Match were significantly related to overall accuracy ($r = -.27$, $p = .03$ and $r = -.40$, $p = .002$, respectively).

It was also possible to use this data to examine the mediating effect of categories on the relationship between training and accuracy. Training was coded as a dichotomous variable as pre-training versus post-training and was correlated with raters' overall accuracy scores ($r = -.27$, $p = .004$). When the effects of the six cognitive category measures were controlled for in a partial correlation, the correlation between training and accuracy was reduced ($r = -.18$, $p = .04$). Although this test was not optimal as the repeated measures scores were used independently in the correlation and hence the sample size was doubled, it provided some means to test for the mediating effect. It appears that cognitive categories have some mediating effect, but there is still a significant direct relationship between training and accuracy. It is reasonable to assume that while cognitive categories do have some effect on rating accuracy, other factors enter into this process.

When considering these results, it is important to remember that the pre-training scores were not a true "control" by which to compare post-training scores; some knowledge about rating may have been gained during the first session prior to assessment of the pre-training accuracy scores. Thus, these relationships may be underestimates of the true effects, if a control group with no prior experience was used.

Conclusion. Taken together, the findings presented here indicate that rater training should be expanded to include

components that concentrate directly on raters' cognitive categories for appraising performance. The present study incorporated only one component of raters' cognitive processes, namely the storing of information into categories. Future research on rater training programs could concentrate on observational skills and recall processes of raters to fully incorporate the cognitive processes of raters into training programs. In addition, most studies investigating the effects of rater training have been lab studies using undergraduate students as raters and not the actual persons who make evaluations of others. Because this study was conducted in a field setting using "real world" people as subjects, it is evident that rater training can be successfully applied in organizational settings.

References

- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. Academy of Management Review, 6, 205-212.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and dairy keeping on psychometric error in ratings. Journal of Applied Psychology, 62, 64-69.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rating errors. Journal of Applied Psychology, 64, 412-421.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 52, 177-193.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process approach. In B. M. Staw & L. L. Cummings (Eds.), Research in organizational behavior (Vol 5, pp. 141-197). Greenwich, CT: JAI Press.

- Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric errors in ratings. Journal of Applied Psychology, 64, 502-508.
- Landy, F. J., & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.
- Ostroff, C., & Ilgen, D. R. (1985). The relationship between cognitive categories of raters and rating accuracy. Manuscript submitted for publication.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.
- Pulakos, E. D. (in press). The interactive effects of format and training on performance appraisal accuracy. Organizational Behavior and Human Decision Processes.
- Smith, P. E., & Kendall, L. M. (1963). Retranslation of expectations: An approach to construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Footnotes

¹Additional data, not relevant to the focus of this study were collected during this phase. The results of these are reported in Ostroff & Ilgen (1985).

²Although the differences in variance suggest a lack of homogeneity of variance across treatment conditions, the proportionality of the cell means and the fact that ANOVAs are quite robust to violations in homogeneity of variance when proportionality exists (Winer, 1971) suggests that the ANOVAs are appropriate analyses.

LIST 1 MANDATORY*

Defense Technical Information Center (12)
ATTN: DTIC DDA-2
Selection & Preliminary Cataloging Section
Cameron Station
Alexandria, VA 22314

Library of Congress
Science and Technology Division
Washington, DC 20540

Office of Naval Research (3)
Code 4420E
800 N. Quincy Street
Arlington, VA 22217

Naval Research Laboratory (6)
Code 2627
Washington, DC 20375

Office of Naval Research
Director, Technology Programs
Code 200
800 N. Quincy Street
Arlington, VA 22217

LIST 2 ONR FIELD

Psychologist
Office of Naval Research
Detachment, Pasadena
1030 East Green Street
Pasadena, CA 91106

LIST 3 OPNAV

Deputy Chief of Naval Operations
(Manpower, Personnel & Training)
Head, Research, Development, and
Studies Branch (OP-115)
1812 Arlington Annex
Washington, DC 20350

Director
Civilian Personnel Division (OP-14)
Department of the Navy
1803 Arlington Annex
Washington, DC 20350

Deputy Chief of Naval Operations
(Manpower, Personnel, & Training)
Director, Human Resource Management
Plans & Policy Branch (OP-150)
Department of the Navy
Washington, DC 20350

LIST 4 NAVMAT & NPRDC

Program Administrator for Manpower,
Personnel, and Training
MAT-0722
800 N. Quincy Street
Arlington, VA 22217

Naval Material Command
Management Training Center
NAVMAT 09M32
Jefferson Plaza, Bldg #2, Rm 150
1421 Jefferson Davis Highway
Arlington, VA 20360

Naval Material Command
Director, Productivity Management
Office
MAT-00K
Crystal Plaza #5, Rm 632
Washington, DC 20360

Naval Personnel R&D Center (4)
Technical Director
Director, Manpower & Personnel
Laboratory, Code 06
Director, System Laboratory, Code 07
Director, Future Technology, Code 41
San Diego, CA 92152

*Number in parentheses is the number of copies to be sent.

Navy Personnel R&D Center
Washington Liaison Office
Ballston Tower #3, Rm 93
Arlington, VA 22217

LIST 5 BUMED

NONE

LIST 6

NAVAL ACADEMY AND NAVAL POSTGRADUATE SCHOOL

Naval Postgraduate School (3)
ATTN: Chairman, Dept. of
Administrative Science
Department of Administrative Sciences
Monterey, CA 93940

U.S. Naval Academy
ATTN: Chairman
Department of Leadership & Law
Stop 7-B
Annapolis, MD 21402

LIST 7 HRM

Officer in Charge
Human Resource Management Division
Naval Air Station
Mayport, FL 32228

Human Resource Management School
Naval Air Station Memphis (96)
Millington, TN 38054

Commanding Officer
Human Resource Management School
Naval Air Station Memphis
Millington, TN 38054

LIST 8 NAVY MISCELLANEOUS

Naval Military Personnel Command (2)
HRM Department (NMPC-6)
Washington, DC 20350

LIST 9 USMC

Headquarters, U.S. Marine Corps
ATTN: Scientific Adviser,
Code RD-1
Washington, DC 20380

LIST 10 OTHER FEDERAL GOVERNMENT

Dr. Brian Usilaner
GAO
Washington, DC 20548

Social and Developmental Psychology
Program
National Science Foundation
Washington, DC 20550

Office of Personnel Management
Office of Planning and Evaluation
Research Management Division
1900 E. Street, NW
Washington, DC 20415

LIST 11 ARMY

Technical Director (3)
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Head, Department of Behavior
Science and Leadership
U.S. Military Academy
New York 10996

LIST 12 AIR FORCE

Air University Library
LSE 76-443
Maxwell AFB, AL 36112

Head, Department of Behavioral
Science and Leadership
U.S. Air Force Academy
Colorado 80840

LIST 13 MISCELLANEOUS

Dr. Eduardo Salas
Human Factors Division
Code 712
Navy Training Systems Center
Department of the Navy
Orlando, FL 32813-7100

LIST 14 CURRENT CONTRACTORS

Dr. Janet L. Barnes-Farrell
Department of Psychology U-20
University of Connecticut
406 Cross Campus Road
Storrs, CT 06268

Jeanne M. Brett
Northwestern University
Graduate School of Management
2001 Sheridan Road
Evanston, IL 60201

Dr. Terry Connolly
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332

Dr. Richard Daft
Texas A&M University
Department of Management
College Station, TX 77843

Dr. Randy Dunham
University of Wisconsin
Graduate School of Business
Madison, WI 53706

Dr. Lawrence R. James
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

Dr. J. Richard Hackman
School of Organization & Management
Box 1A
Yale University
New Haven, CT 06520

Dr. Frank J. Landy
Department of Psychology
Pennsylvania State University
450 Moore Bldg.
University Park, PA 16802

Dr. Bibb Latane
University of North Carolina
at Chapel Hill
Manning Hall 026A
Chapel Hill, NC 27514

Dr. Edward E. Lawler III
Graduate School of Business
University of Southern California
Los Angeles, CA 90007

Dr. William H. Mobley
College of Business Administration
Texas A&M University
College Station, TX 77843

Dr. Thomas M. Ostrom
The Ohio State University
Department of Psychology
404C West 17th Avenue
Columbus, OH 43210

Dr. Robert Rice
Department of Psychology
SUNY - Buffalo
Buffalo, NY 14226

Dr. Benjamin Schneider
Department of Psychology
University of Maryland
College Park, MD 20742

Dr. H. Wallace Sinaiko
Program Director, Manpower Research
and Advisory Services
Smithsonian Institution
801 N. Pitt Street, Suite 120
Alexandria, VA 22314

Dr. Richard M. Steers
Graduate School of Management
University of Oregon
Eugene, OR 97403

Dr. Harry C. Triandis
Department of Psychology
University of Illinois
Champaign, IL 61820

Dr. Anne S. Tsui
Duke University
The Fuqua School of Business
Durham, NC 27706

Andrew H. Van de Ven
University of Minnesota
Office of Research Administration
1919 University Avenue
St. Paul, MN 55104

END

FILMED

1-86

DTIC